厚生労働科学研究研究費補助金

健康科学総合研究事業

地理及び社会状況を加味した地域分析方法の開発に関する研究

平成１６年度　総括・分担研究報告書

主任研究者　浅見　泰司

平成１７（２００５）年　４月

目　　次

厚生労働科学研究費補助金（健康科学総合研究事業）
総括研究報告書

地理及び社会状況を加味した地域分析方法の開発に関する研究

主任研究者　浅見 泰司　東京大学空間情報科学研究センター教授

研究要旨

本研究では，健康危機情報などをリアルタイムで空間情報化し，特異性検出の分析を行うことで行政対応を支援する総合的な地域診断システムの開発を行うべく，(1)有効かつ迅速なデータの取得方法についての検討，(2)非定型書式データから有効なデータを取得する技術の開発，(3)リアルタイムであがってくる健康関連基礎情報をデータ書式変換などが不要で簡易に地図表示・加工を実現するために，G-XML に適合した形で空間情報を流通できる WebGIS の構築技術の開発，(4)その情報を用いて，短期間に有効な意思決定ができるための症候の時間・空間的増加あるいは集積性を評価するための統計モデル開発，(5)それを補助するための空間マイニング・空間推論などの空間解析支援ツールの技術開発，および，(6)健康危機管理政策上に有益な知見を得るための当システムの有効性の検証を行った．

分担研究者

丹後 俊郎
国立保健医療科学院
技術評価部部長

郡山 一明
救急救命九州研修所教授

有川 正俊
東京大学空間情報科学
研究センター助教授

A．研究目的

現在の地域保健行政において，現状把握，問題抽出，原因分析等は，社会に存在する情報の一部の情報である保健・医療・福祉に関した統計調査等を画一的側面から分析しているため，正確な現状把握及び本質的原因の解明を行うことは困難である．また，現状実施されている統計調査はリアルタイムの報告ではない上に，市町村等空間的に大きな集計単位の統計数値となっているため，社会の複雑化，高速化に対応して迅速に社会の状態を捉えた現状把握やCommunity単位やFamily単位での詳細な状況分析，問題の早期発見をも困難としている．

本研究では，この状況を打開し，迅速かつ効率的に現状把握等を実施するために，因子間の相互関係の解明，知的情報処理技術による時空間データを含む社会情報の収集・入力技術の確立，それに合った空間解析・統計モデルの開発をおこなうものとする．

B．研究方法

健康危機などに際して，リアルタイムで精密な行政対応を支援する総合的な地域診断システムの開発を目的として，以下の通り研究を実施した．

1) さまざまなドキュメント(EXCEL, HTML など)の中から自動的に住所を抽出し，瞬時に緯度経度を導出して，地図の上で空間的分布を閲覧できるソフトウェアシステム「空間ドキュメント管理システム(SDMS: Spatial Document Management System)」の基本設計および実現可能性に関して検討をおこなった．また，上記の基

本設計にしたがったプロトタイプの開発をおこなった．1)のシステムによって，健康危機イベントの場所が特定され，直感的な状況把握が可能となった主題図は，2) WebGIS を通じて，リアルタイムで利用者に周知され共有される．

そこで，上記の空間情報システムと将来的に統合し得る，地域状況の新たな分析手法として，3) 疾病などの集積性の検出およびその集積地域を同定する検定法として，従来よく用いられる Kulldorff の方法を改良した新たな手法である Flexible scan 法の開発，ならびに 4) 感染症の流行過程における空間的パターンの発見を目的とした，疾病空間情報の早期の把握および将来的予測を可能にするための空間解析手法の開発に関する研究を行った．

また，5) 上記で述べたシステムが稼働する際の，地域サーベイランスとしての意義を検証し，かつ，早急に対策が望まれる状況について実際に分析することによって，健康危機管理政策上に有益な知見を得るための研究をおこなった．そして最後に，以上によって得られる本研究の成果についての総括をおこなった．

（倫理面への配慮）

本研究においては，原則として公開されたデータを用いて技術開発を行なう．また，プライバシーにかかわる個人情報を扱う場合は，個人情報の保持・公開には十分留意すると共に，疫学倫理指針に基づき研究を行なうこととする．

C．研究成果

本研究によって得られた研究成果は主に以下の通りである．（項目の数字は「B．研究方法」のものにしたがう）

1)「空間ドキュメント管理システムの設計と開発に関する研究」

空間ドキュメント管理システムのシステム設計を行い，また，現在の実現環境下での実現性に関して整理することができた．システム設計にしたがいプロトタイプを開発した．

3)「集積地域同定のための新たな検定手法の開発に関する研究」

現在よく用いられている Kulldorff の方法を改良する形で新たな方法 Flexible scan 法を開発した．この手法によって，様々な形状の疾病集積地域を同定できるようになった．

4)「時間的推移に伴うパターン発見のための空間解析手法の開発に関する研究」

最尤法を用いた定点報告数の母数についての推計方法の提案，および感染症の時間的変化にともなう感染状況をあらわす空間モデルの定式化をおこなうとともに，1)のシステムを実際に用いて，住所情報から位置情報を取得し，部分的に GIS への統合を実現した．

5)「健康危機管理における地域サーベイランスの意義」

本研究で開発されたシステムが，症候群サーベイランスとして有効であり得るかを検証するために，学級閉鎖状況と定点観測データとの関連性，および特定物質の地理的な差異について示すとともに，地図を用いた社会状況の地域住民への伝達方法について考察した．さらに，健康危機管理に関する「地理及び社会状況を加味した地域分析方法の開発」システム構築についてガイドラインを作成した．

D．考察

「空間ドキュメント管理システム(SDMS)」は，知的情報処理によって，比較的自由な書式でありながら，汎用性に優れた G-XML に適合した形で空間情報を流通できるシステムが確立できるため，健康情報の収集から空間データ化するまでの時間や手間を大きく省くことができることのみならず，今後モバイル環境での位置情報利用などで活用できる可能性がある．また，一般ユーザでも抵抗なく利用できることか

ら，WebGIS 上での健康危機情報シェアリングに直結できる．

当システム内に含まれる，ジオコーディングと半構造化手法を用いることによって自動的に得られる座標値を用いて，「時間的推移に伴うパターン発見のための空間解析手法の開発に関する研究」では，GIS 上で学級閉鎖状況を把握することが可能な地図を作成するなど，今後の GIS への統合を見据え，その有効性を実証した．同じく，「健康危機管理における地域サーベイランスの意義」でも，健康危機管理データの地図上への表示や，AED 配置や災害要援護者の事前登録などへの応用について，当システムを利用することで，分析がより有利になることが期待できる．

将来的にこのシステムに統合する「集積地域同定のための新たな検定手法の開発に関する研究」では，任意の形状の集積地域を同定することが可能な Flexible scan 法が提案され，その手法としての有効性が Kulldorff の方法との比較によって確かめられたとともに，モンテカルロシミュレーションを用いたときの実際の計算の実行速度についても検討した．同様に計算速度に関しては，「時間的推移に伴うパターン発見のための空間解析手法の開発に関する研究」のなかでも，空間モデルのパラメータを推定する際に，非線形最小二乗推定法のアルゴリズムにサブ空間 trust region 法を用い，適当な制約条件を付することで収束可能であった．しかし，全医療圏を対象としたときの計算時間は短いとはいえず，得られるパラメータ値の安定性とともに，モデルの柔軟性についても今後検討する必要があると考えられる．

また，「健康危機管理における地域サーベイランスの意義」では，実際の健康危機管理上の視点から，データの集積・蓄積についての概念をはじめとしたガイドラインが提案され，当システム構築の今後の指針として十分検討されるべきである．

E．結論

本研究で開発されたシステムは，従来の地理情報システムでは取り扱えなかった，人間活動で日常的に利用する位置を表現するテキスト情報を，知的言語処理と住所マッチング技術を用いて自動的に空間データ化するシステムを構築しようというユニークなもので，今後，位置情報の利用の拡大を導くものである．

症候群サーベイランスとして，分析上必要となる各種のデータは，統計解析や空間解析を行うに可能な状態で整備されていないか，あるいは，速やかに公開されないものであるものが少なくない．本システムでは，健康情報の収集から空間データ化するまでの時間や手間を大きく省くことができ，このような状況を補完するものとして極めて実用性が高い．

様々な形状の集積地域を同定できる検定法である Flexible scan 法，および時系列データにおけるパターン発見を目的とした空間解析手法は，地理及び社会状況を加味した地域分析方法として本システムと連動することで，健康被害症候の特異性を早期に検出し，またその原因や健康被害の空間的拡散の特色を適切に推測する次世代型の地域診断システムとして，健康で安全な社会の構築に貢献できる．

F．研究発表

1．論文発表

[1]片岡裕介，及川清昭，浅見泰司(2004)「迷惑施設の立地適性に関する数理的考察」『都市計画論文集』39-3, 829-834.

[2]Yasushi Asami, Ayse Sema Kubat, Istek Cihangir (2004) “Application of GIS to Network Analysis: Characterization of Traditional Turkish Urban Street Network” Atsuyuki Okabe (ed.) Islamic Area Studies with Geographical Information Systems, Routledge-Curzon, Taylor &Francis Group, London, pp.187-206.

[3]Yasushi Asami, Ayse Sema Kubat, Kensuke

Kitagawa, Shinichi Iida (2004) "A Three-Dimensional Analysis of the Street Network in Istanbul: An Extension of Space Syntax Using GIS" Atsuyuki Okabe (ed.) Islamic Area Studies with Geographical Information Systems, Routledge- Curzon, Taylor&Francis Group,London, pp.207-220.
[4]浅見泰司(2004)「新技術と都市の変化」『都市計画』249, 5-9.

2. 学会発表

[1]Tango T. and Takahashi K. A Flexible Scan Statistic for Detecting Arbitrarily Shaped Clusters, Joint Statistical Meetings 2004, Toronto, Canada, Abstracts p.14.
[2]Takahashi K. and Tango T. How to Evaluate Tests for Identifying Spatial Clusters, Joint Statistical Meetings 2004, Toronto, Canada, Abstracts p.14.
[3]高橋邦彦，丹後俊郎．平面領域同定の検定における評価指標，2004年度統計関連学会連合大会，富士大学，岩手，講演報告集 p.288.
[4]丹後俊郎，高橋邦彦，横山徹爾．疾病の集積地域同定のための新しい検定法，第15回日本疫学会学術総会，滋賀，講演集 p.180.
[5]横山徹爾，高橋邦彦，丹後俊郎．Flexible scan法を用いた疾病集積地域同定ソフトウェアの開発と応用例，第15回日本疫学会学術総会，滋賀，講演集 p.181.
[6]Kouzou Noaki, Masatoshi Arikawa: Geocoding Natural Route Descriptions using Sidewalk Network Databases, International Workshop on Challenges in Web Information Retrieval and Integration (WIRI2005), IEEE, April 8-9, 2005, NII, to be published from IEEE Computer Science Press.
[7]野秋浩三, 有川正俊 : A Method for Parsing Route Descriptions using Sidewalk Network Databases, 電子情報通信学会データ工学研究専門委員会, 第16回データ工学ワークショップ(DEWS2005)講演論文集, 3A-i10,2005年2月28日-3月2日, Web 掲載.
[8]林 徹，有川正俊：地図Blogを対象とした幾何形状を用いた間接トラックバック手法，電子情報通信学会データ工学研究専門委員会, 第16回データ工学ワークショップ(DEWS2005)講演論文集, 3A-i4, 2005年2月28日-3月2日，Web掲載.
[9]Engindeniz, E. and Y. Asami (2004) "Understanding of City Structure: A Cognitive Map Analysis of Central Tokyo" Proceedings of the 6th International Symposium for Environment Bahavior Studies, Baihua Literature and Art Publishing House, pp.73-84.

3. ソフトウェア

Takahashi K, Yokoyama T and Tango T. FleXScan v1.1: Software for the Flexible Scan Statistic. National Institute of Public Health, Japan, 2005.

G. 知的財産権の出願・登録状況

1. 特許取得	該当なし
2. 実用新案登録	該当なし
3. その他	該当なし

厚生労働科学研究費補助金（健康科学総合研究事業）
分担研究報告書

空間ドキュメント管理システムの設計と開発に関する研究

分担研究者　有川 正俊　東京大学空間情報科学研究センター助教授
協力研究者　白石　陽　東京大学空間情報科学研究センター研究機関研究員

研究要旨
さまざまなドキュメント（EXCEL, HTML など）の中から自動的に住所を抽出し，緯度経度を導出して，地図の上で空間的分布を閲覧できるソフトウェアシステム「空間ドキュメント管理システム（SDMS: Spatial Document Management System)」の基本設計および実現可能性に関して検討を行った．また，この基本設計にしたがってプロトタイプの開発を行った．

A．研究目的

多くのドキュメントデータには，そのデータが作成された場所や著者の住所，あるいはある場所を参照するといった，実世界の位置の情報が含まれている．このような多様なドキュメントデータを実世界の位置で検索・管理することは，情報の活用可能性を広げ，利用を高度化させる．一般に，位置情報というと緯度経度で表される２次元座標値が想定されるだろう．２次元座標値のように位置を数値で表したものを*直接位置情報*と呼ぶ．直接位置情報を利用する代表的な応用例としては，GIS(Geographic Information Systems) や GPS(Global Positioning System)が挙げられるが，直接位置情報が利用されているのは，特定の目的に作られた専門性の高いデータだけである．これに対し，住所や地名のように位置の情報を表しているものの，直接地図上に射影できない記述を*間接位置情報*と呼ぶ．間接位置情報を含むドキュメントは一般ドキュメントデータにも多数存在する．これらの間接位置情報を直接位置情報，つまり(x, y)へと変換できれば，ドキュメントデータを地理空間に射影することができ，多様な検索や情報の構造化，そして空間解析が可能となる．

間接位置情報を直接位置情報へ変換する手法として，欧米を中心に従来よりジオコーディング(Geocoding)が利用されている．昨今のモバイルコンピューティング環境の普及に伴い，位置に基づく検索・整理や，情報発信が今後ますます重要になると考えられるため，ジオコーディング手法を利用することで，メディアの種類を超えて，位置に依存したさまざまなアプリケーションが一般ユーザにも使えるようになるのが理想と言える．このようなインフラが整備されれば，現実世界とのインタラクションのある空間情報を日常的に利用できるようになるだろう．

われわれは，本研究補助金を受ける以前より，ジオコーディング・エンジンの開発などを行い研究成果を上げてきた．本研究では，すでに開発したエンジンを応用してさまざまな一般ドキュメントデータを空間情報化し，位置によりアクセスする現実的な枠組みを体系化した．さらに，この枠組みを利用した「空間ドキュメント管理システム」を実際に構築することにより，提案した枠組みが実用に耐えうるかを示した．

B．研究方法

本研究の対象とするドキュメントデータを，空間データとして利用する観点から分類する．最も代表的な空間データを扱う情報システムである GIS で利用可能なデータには，地理データ(Geographic Data)と地理参照データ(Geo-referenced Data)がある．地理データは道路形状や行政界などの幾何

的な情報を中心としたものである．地理参照データは，顧客データや道路交通量などの定量データが中心だが，IDなどによって地理データにリンクすることができるデータである．地理データも地理参照データも特定目的用に多くの費用をかけて作成されるもので，一般ユーザが日常的に利用するものではない．

さて，日常生活で利用されるドキュメントに含まれる情報には，待ち合わせ場所や宿泊先など，住所や地名を含むものが多い．このような「空間的な位置情報を含むデータ」を「空間データ(Spatial Data)」と定義する．空間データには，「○△町□番地で火災発生」「震源地は××沖 50km」や「○○駅前の△ラーメンはおいしい」といった自然言語で記述された文章や，略地図，事故現場を写すニュース映像なども含まれる．

このような高級な表現は人間にとっては有用だが，そのままではコンピュータには理解できないため利用できない．そこで，XML などの半構造化表現を利用したドキュメント記述を用いて，表現の曖昧さを解消する手法が注目されている．たとえば「○△町□番地で火災発生」というデータを「<spatial information><location>○△町□番地</location>で<event>火災発生</event></spatial information>」のように記述すれば，コンピュータにとって格段に理解しやすくなる．

ドキュメントデータを空間データとして分類すると，上述のように構造化のレベルによって3段階に分類することができる．まず自然言語や画像などの生データを「非構造データ(Non-structured Data)」，XMLのような構造化ドキュメント表現を利用したデータを「半構造化データ(Semi-structured Data)」，そして地理データや表形式データのように特定のフォーマットに従ったデータを「構造化データ(Full-structured Data)」と分類する．これと直交する基準として，地理データのように位置を座標値で表現した直接位置情報データ(Directly location referenced data)と，位置を住所や地名で表現した間接位置情報データ(Indirectly location referenced data)に分けることができる．以上の組み合わせにより，空間データを図1のように6種類に分類することができる．以下本稿ではそれぞれの頭文字を用いて，構造化－直接位置情報データを F-D データ（Full-structured, Directly location referenced data），非構造化－間接位置情報データを N-I データ(Non-structured, Indirectly location referenced data)のように表記する．

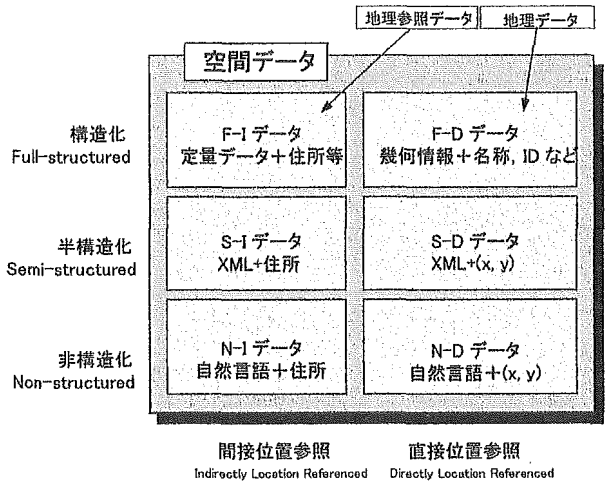


図 1：空間データの分類

本研究の目的である空間ドキュメントデータの高度利用を実現するためのシステムとして，空間ドキュメント管理システム(SDMS: Spatial Document Management System)を実装・開発し，有効性を確認した．本システムでは，文章で記述されているドキュメントデータであれば(すなわち，画像や音声のようなデータは除く)，前節で分類した6種類の空間ドキュメントをすべて空間情報として利用することができる．たとえばレストランの情報であれば，ワープロで作成されたチラシや Web ページのような

N-I ドキュメントデータもそのまま保存し，含まれている住所の情報を元に地図上で検索，閲覧することができる．以下，空間ドキュメント管理システムで利用する２種類の変換エンジンについて説明し，次に空間ドキュメント管理システムについて説明する．

(B-1)　ジオコーディング・エンジン

空間ドキュメント管理システムでは，間接位置情報を抽出して直接位置情報に変換するため，ジオコーディングを行う必要がある．ジオコーディングは，住所や地名文字列を解釈し，対応する位置の座標値（たとえば緯度経度）に変換する手法の総称である．欧米では GIS の基本機能として広く利用されているが，日本では，単語の間に空白やカンマなどのデリミタが存在しないため分かち書き処理を行う必要があることや，京都市の通名に代表されるように複数の住所体系が混在していることなどが障害となり，あまり普及していない．特に一般ドキュメントデータに含まれる住所などの記述は，読み手に理解できればよいという条件で記述されているため，都道府県名や市町村名が省略されているなど曖昧な記述が多い．われわれは，これらの曖昧な間接位置情報をロバストかつ高速にジオコーディングするため，日本の住所体系に適したジオコーディングアルゴリズムを開発し，クライアント・サーバエンジン『SPAT』として実装した．空間ドキュメント管理システムでも SPAT を呼び出してジオコーディングを行う．

ジオコーディングにより，S-I データは S-D データに，F-I データは F-D データに変換される．

(B-2)　半構造化エンジン

非構造化データには，間接位置情報がどこに記述されているかという情報が含まれていない．そこで，文章をパース(parse)して，間接位置情報の可能性がある単語列を順番にジオコーディングするという処理を行う．ジオコーディングの結果，対応する緯度経度が得られれば間接位置情報であったことが分かると同時に，直接位置情報に変換することができる(対応する緯度経度が得られなかった場合は間接位置情報ではなかったと判断し，次の単語列に移る)．また，直接位置情報の可能性がある単語列も抽出する．

さて，元の非構造ドキュメントデータに含まれる間接・直接位置情報が抽出された時，その部分を XML-like なタグでマークアップすると，非構造化データを半構造化データに変換することができる．そこで，この処理を「半構造化(semi-structuralize)」と呼ぶ．実際には，同時にタグの属性情報として直接位置情報を挿入するため，半構造化とジオコーディングが行われる．すなわち，N-I データと N-D データが S-D データに変換される．

(B-3) 空間ドキュメント管理システムの仕組み

ジオコーディング・エンジンおよび半構造化エンジンを利用することで，6 種類に

図 2：SDMS のプロトタイプ画面

分類された空間ドキュメントデータはすべて，S-Dデータまたは F-Dデータに変換できる．一般に S-D データはレストラン情報のように地図上の点として表される情報，いわゆる POI(Point of Interest)とみなすことができ（道路渋滞情報のように線で表されるべき情報もある)，地図に表すことができる．F-D データはそのまま地図上に表示することができるため，6 種類の分類すべてが地図上に示せることになる．

そこで，変換された S-D データ及び F-D データを効率良く管理，検索する仕組みを開発すれば，6 種類の空間ドキュメントデータを地図上で管理できる新しい情報システムを構築することができる．この空間情報システムを「空間ドキュメント管理システム」と呼ぶ．図 2 は SDMS のプロトタイプシステム画面例である．

C．研究成果

空間ドキュメント管理システムのシステム設計を行い，また，現在の実現環境下での実現性に関して整理することができた．システム設計にしたがいプロトタイプを開発した．開発したプロトタイプは，一般公開可能であり，テスト的に利用していただくことができる．実装に関しては，確実な部分を行ったが，同時に，次の研究展開を見据えて，高度なアドレスジオコーディングならびに Blog の地図利用への拡張に関しての検討も行った．

D．考察

日常生活で利用される住所などの間接位置情報を含んだドキュメントデータから，ジオコーディングと半構造化手法を用いることで，自動的に緯度経度などの座標値を算出し地図上にリンクする空間ドキュメント管理システムを構築した．このシステムでは，空間情報をドキュメント形式のまま管理し利用できる点に特徴があり，今後モバイル環境での位置情報利用などで活用できる可能性がある．コンピュータには理解が困難なドキュメント形式の情報を扱うため，従来の GIS に比べて登録や検索に時間がかかるが，データの作成や再利用が容易なこと，専用ソフトや知識が不要なことといった利点があり，一般ユーザでも抵抗なく利用できる．

E．結論

従来の地理情報システムで対象としていたデータは，緯度経度などの直接位置情報データだけと言って良い．しかしながら，直接位置情報データは無いが，住所などの間接位置情報データとして位置データが表現されているドキュメントは膨大にある．本研究は，そのような従来の地理情報システムでは取り扱えなかった，人間活動で日常的に利用する位置を表現するテキスト情報を，自然言語処理や情報検索の技術を利用して，利用できるようにしたという点で，今後，位置情報の利用の拡大を導くものである．

今年度は，空間ドキュメント管理システムの実現可能性を中心にシステム設計を行い，プロトタイプシステムを作成した．実際に使えるようにするためには，ユーザビリティに対する改善が必要である．また，ジオコーディングに関して，現在，住所から緯度経度という点位置情報に変換しているが，より正確には，面位置情報や線位置情報に変換する必要があり，これらの拡張を次年度に行う予定である．すでに開発したプロトタイプは，試験版として一般公開する予定である．

本研究を遂行するにあたり，アドレスマッチングのエンジン部分を本システム向けに改良を加えて利用させていただきました東京大学生産技術研究所の相良毅助手に感謝いたします．また，相良毅助手には，空間ドキュメント管理システムの設計の際に多くの有意義なアドバイスとコメントをいただきました．

F. 研究発表

1. 論文発表

なし.

2. 学会発表

[1] Kouzou Noaki, Masatoshi Arikawa: Geocoding Natural Route Descriptions using Sidewalk Network Databases, International Workshop on Challenges in Web Information Retrieval and Integration (WIRI2005), IEEE, April 8-9, 2005, NII, to be published from IEEE Computer Science Press.

[2] 野秋浩三, 有川正俊：A Method for Parsing Route Descriptions using Sidewalk Network Databases, 電子情報通信学会データ工学研究専門委員会, 第16回データ工学ワークショップ(DEWS2005)講演論文集, 3A-i10, 2005年2月28日-3月2日, Web 掲載.

[3] 林 徹, 有川正俊：地図Blogを対象とした幾何形状を用いた間接トラックバック手法, 電子情報通信学会データ工学研究専門委員会, 第16回データ工学ワークショップ(DEWS2005)講演論文集, 3A-i4, 2005年2月28日-3月2日, Web掲載.

G. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

厚生労働科学研究費補助金（健康科学総合研究事業）
分担研究報告書

集積地域同定のための新たな検定手法の開発に関する研究

分担研究者　丹後 俊郎　国立保健医療科学院技術評価部部長
協力研究者　高橋 邦彦　国立保健医療科学院技術評価部研究員

研究要旨
地域診断システムにおけるバイオ・サーベイランスの方法として，疾病などの集積性の検出およびその集積地域を同定する検定法について研究を行った．本研究では，従来よく用いられるKulldorffの方法を改良する形で，新たな手法 Flexible scan 法を開発した．この方法により，従来法では同定できないような様々な形状の集積地域を同定することができるようになり，実際，シミュレーションによる検討でも，より真の集積地域に近い領域をうまく同定できる様子が確認された．

A．研究目的

地域の慢性的な健康問題や健康被害の特異性を早期に検出し注意喚起を行う地域診断システムにおいて，症候の時間・空間的特異性を評価するための統計モデル開発は重要である．特に，統計的に有意な疾病の空間集積性検出(spatial clutering)のための検定法はバイオ・サーベイランスの方法として用いられている．中でも有意な集積地域を同定する検定法(cluster detection test)は本研究課題におけるバイオ・サーベランスの早期対策を立てる上で大変有用である．本研究では，世界的に広く用いられている Kulldorff の方法を改良する形で新たな検定手法 Flexible scan 法を提案する．

B．研究方法

集積性の検定の中でも，実際の集積地域を同定することも含んだ検定(cluster detection test)では，Turnbull らの方法(1990)，Besag and Newell の方法(1991)，Tango の方法(2000)などいくつか提案されて実際利用されている．その中でも，Kulldorff の方法(1995)は，その簡便さやアプリケーションソフトとして普及している点から最も広く用いられてる．Kulldorff の方法は Scan 統計量に基づき，最大尤度比を用いる検定であるが，その scan 統計量の設定の問題から，円状に近い地域(cluster)しか同定できない性質がある．しかし，実際の疾病の集積地域としては，河川や道路に沿った地域など様々な形状の地域が考えられ，必ずしも円状とは限らない．そこで我々は，任意の形状の集積地域を同定できるような scan 統計量として flexible scan 統計量を定義し，それに基づく Flexible scan 法を提案する．ここでは特に疾病による死亡状況の集積性を例として説明する．

まず，対象地域が m 個の地区（市区町村など）に分割されているとし，i 地区の死亡数 N_i が確率変数として，期待値 $p_i\xi_i$ をもつポアソン分布 $Po(p_i\xi_i)$ に従うというモデルを考える．ただし，ξ_i は i 地区における期待死亡数とし，p_i は i 地区のリスク（標準化死亡比 SMR）とする．また，実際観測された死亡数を n_i とする．

このとき，対象地域内の連結した地区によるクラスター Zを考え，Z 内での観測死亡数 $N(Z)$，期待死亡数 $\xi(Z)$ に対し

$$E(N(Z)) > \xi(Z)$$

となるクラスター Zが，疾病による死亡が集積している地域と定義する．つまり，対

象地域において疾病の集積がない場合

$$E(N(Z)) = \xi(Z) \quad \text{for all } Z(\in \mathcal{Z})$$

となる．そこで，Kulldorffは上記の設定において，あるクラスター$Z(\in \mathcal{Z})$をとったとき，

$$p_i = p\ (i \in Z);\quad p_i = q\ (i \notin Z)$$

というモデルを考え，

帰無仮説$H_0 : p = q$，対立仮説$H_1 : p > q$

で，$\frac{n(Z)}{\xi(Z)} > \frac{n(Z^c)}{\xi(Z^c)}$のもとでの最大尤度比

$$\lambda = \sup_{Z \in \mathcal{Z}} \left(\frac{n(Z)}{\xi(Z)}\right)^{n(Z)} \left(\frac{n(Z^c)}{\xi(Z^c)}\right)^{n(Z^c)}$$

をもつクラスターZ^*を見つけ出すことを提案した．その際，考え得るクラスターZ全てにおいて尤度比を計算しその最大値を探す(scan)のではなく，区域iを中心に半径rの円を描き，その円に含まれる領域をZとした．このrを0から予め設定された上限まで連続的に変化させ，さらにそれを全てのiについて行うことで様々な領域を取ることができる．このようにして得られる全ての領域Zの集合として$\mathcal{Z}$を定めた．この方法は非常に明解かつ簡便であり，実際多くの疫学研究の場面で利用されている．しかしながら，その$\mathcal{Z}$の定め方から円状に近い平面領域の同定には優れているが，円状でない領域の場合にはうまく同定されないことになる．そこで我々では以下のような$\mathcal{Z}$を定める方法 Flexible scan 法を提案する．

1. ある区域iからの距離が短い順に$k-1$個の区域を取り出す．
2. 取り出した$k-1$個の区域とiを合わせたk個の中から長さ$l(\leq k)$個の組み合わせを考え，その中でiを含み，さらに全てが連結している組み合わせをZとする．
3. lを1からkまで変化させ，全てのZを取り出す．
4. 以上の手順を各iについて行い，その全てのZを要素とする集合を$\mathcal{Z}$とする．

この方法により，Kulldorffの方法より計算する時間は長くなるが，様々な形状のクラスターを同定することができる．

C．研究成果

実際に Flexible scan 法を疾病集積性の検定に適用して，Kulldorffの方法と比較を行った．ここでは東京都と神奈川県を 113 市区町村単位に分けた空間を対象とし，ホットスポットモデル(p=3q)の下でいくつかのシミュレーションによって検討を行った．その際，従来の検出力だけではなく，より詳細に同定の状況がわかる bivariate power distribution(Tango & Takahashi)によって，どの程度正確に集積地域が同定できるか評価を行った．その結果，以下のような特徴を見ることができた．

- Kulldorff の方法は円状の地域はうまく同定できる．
- 円状・非円状とも，Kulldorff の方法では，真ではない余計な地域を含んでしまう広めの地域を同定してしまう傾向があった．
- Flexible scan 法は，真に近い地域をうまく同定できている傾向が見られた．

D．考察

本研究で考えている Cluster detection test では，対象地域に特定の疾病などの集積があるかどうかを判断するのみならず，集積がある場合，その集積地域がどこかを同定することが求められている．従来よく用いられている Kulldorff の方法は，円状のwindowを用いたscan法であるため，円状に近い領域の同定にはすぐれているが，

それ以外の複雑な形の領域の同定はできない. このことは，一般の検出力，つまり“有意な集積はない”という帰無仮説を棄却する確率だけでは観察できないものであり，我々の行った bivariate power distribution を用いることにより，その同定の様子をつかむことができた．本研究で提案する Flexible scan 法は，この Kulldorff の方法を改良する形で，任意の形状の集積地域を同定することができ，bivariate power distribution による Kulldorff の方法との比較でもより真に近い領域を安定して同定していることが確かめられた.

実際の計算ではコンピュータープログラムによるモンテカルロシミュレーションを利用することになり，実用上の観点からその実行速度も大変重要であると考えられる. もちろん Kulldorff の方法の実行速度はかなり短く済むが，Flexible scan 法でもアルゴリズムの改良などを行い，十分に利用できる範囲の実行速度で結果を得ることができた．先の東京，神奈川地区の 113 市区町村のシミュレーションでは，Windows XP, CPU pentium 4, 3.2GHz で，scan する限界値を中心から 15 地区までとした場合で 14 秒，20 地区までにすると 379 秒で終えることができた．この scan する限界値の値の決め方は，対象地域や検討する問題によって考える必要があるが，113 市区町村に対し，最大 15 地区～20 地区の範囲で集積地域を探すことは，全地区数の 1 割～2 割程度であり，ある意味リーズナブルであると考えられ，逆に，あまりにこの最大地区数を伸ばすと，観測数がゼロのような地域も含まれてしまうという問題が生じる可能性がある.

E. 結論

本研究で提案された Flexible scan 法は様々な形状の集積地域を同定できる検定法として，従来の Kulldorff の方法よりも優れているものといえる．今後はこれを平面空間上だけでなく，時間も考慮した集積を検出する方法に拡張する予定である．この手法を利用することによって，地理及び社会状況を加味した地域分析方法として，地域の慢性的な健康問題や健康被害の特異性を早期に検出注意喚起を行う地域診断システムに利用でき，健康を脅かす事象（症候あるいは疾病）の発生を早期に予知するバイオ・サーベイランスの検討に有効な手法になると考えられる.

なお，この Flexible scan 法を実際に簡単に利用できるよう，Windows 上で動作するアプリケーションプログラム“FleXScan”を作成し，現在，国立保健医療科学院技術評価部のホームページ (http://www.niph.go.jp/soshiki/gijutsu/) で無料配布中である.

F. 研究発表

1. 学会発表

Tango T. and Takahashi K. A Flexible Scan Statistic for Detecting Arbitrarily Shaped Clusters, Joint Statistical Meetings 2004, Toronto, Canada, Abstracts p.14.

Takahashi K. and Tango T. How to Evaluate Tests for Identifying Spatial Clusters, Joint Statistical Meetings 2004, Toronto, Canada, Abstracts p.14.

高橋邦彦，丹後俊郎．平面領域同定の検定における評価指標，2004 年度統計関連学会連合大会，富士大学，岩手，講演報告集 p.288.

丹後俊郎，高橋邦彦，横山徹爾．疾病の集積地域同定のための新しい検定法，第 15 回 日本疫学会学術総会，滋賀，講演集 p.180.

横山徹爾，高橋邦彦，丹後俊郎．Flexible scan 法を用いた疾病集積地域同定ソフトウェアの開発と応用例，第 15 回 日本疫学

会学術総会，滋賀，講演集 p.181.

2. ソフトウェア
Takahashi K, Yokoyama T and Tango T. FleXScan v1.1: Software for the Flexible Scan Statistic. National Institute of Public Health, Japan, 2005.

G．知的財産権の出願・登録状況
なし

表：あるホットスポット(地区数 4)を真の集積地域とした場合の Flexible scan 法と Kulldorff の方法による同定の様子を示した bivariate power distribution. 縦は同定された領域の地区数，横は同定された領域に含まれるホットスポット内の地区数．どちらも scan する最大地区数は中心から 15.

Flexible

Length l	Include s hot-spot regions					Total
	0	1	2	3	4	
1	0	0				0
2	0	0	0			0
3	0	0	0	0		0
4	0	0	0	0	127	127
5	1	0	0	0	157	158
6	0	0	0	0	205	205
7	0	0	0	2	198	200
8	0	0	0	1	151	152
9	0	0	0	5	85	90
10	0	0	0	1	24	25
11	0	0	0	0	17	17
12	0	0	0	0	5	5
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
Total	1	0	0	9	969	979

usual power=0.979

Kulldorff

Length l	Include s hot-spot regions					Total
	0	1	2	3	4	
1	0	0				0
2	0	0	0			0
3	0	0	0	523		523
4	0	0	0	65	0	65
5	0	0	0	23	0	23
6	0	0	0	7	66	73
7	0	0	0	0	15	15
8	0	0	0	0	32	32
9	0	0	0	1	15	16
10	0	0	0	0	7	7
11	0	0	0	2	3	5
12	0	0	0	2	63	65
13	0	0	0	0	96	96
14	0	0	0	0	30	30
15	0	0	0	0	22	22
Total	0	0	0	623	349	972

usual power=0.972

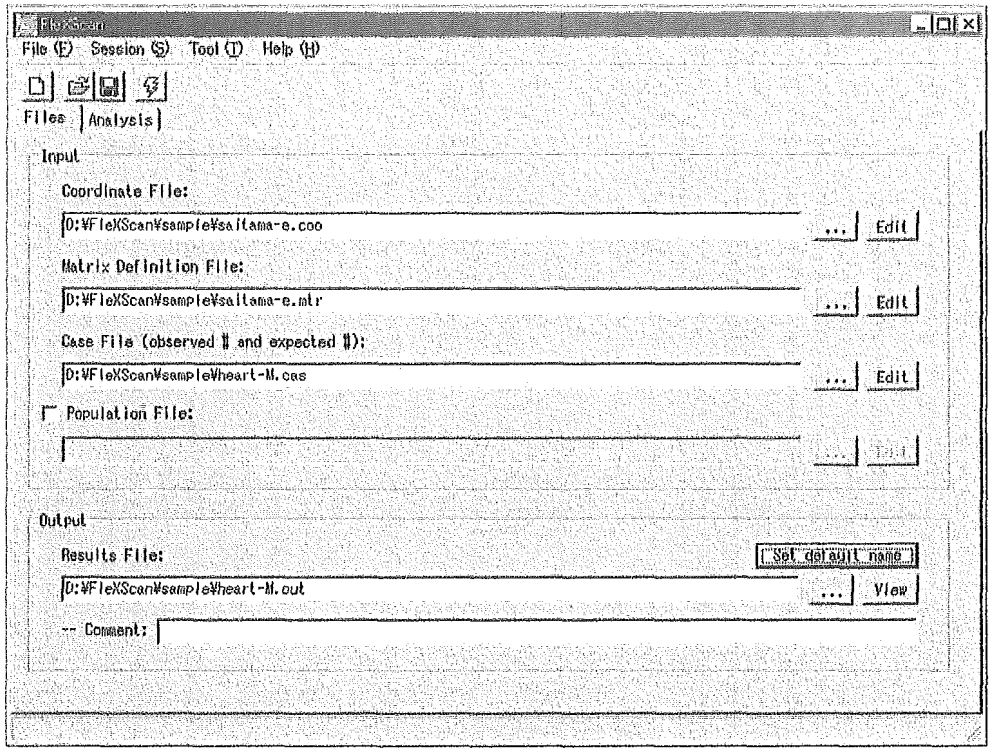
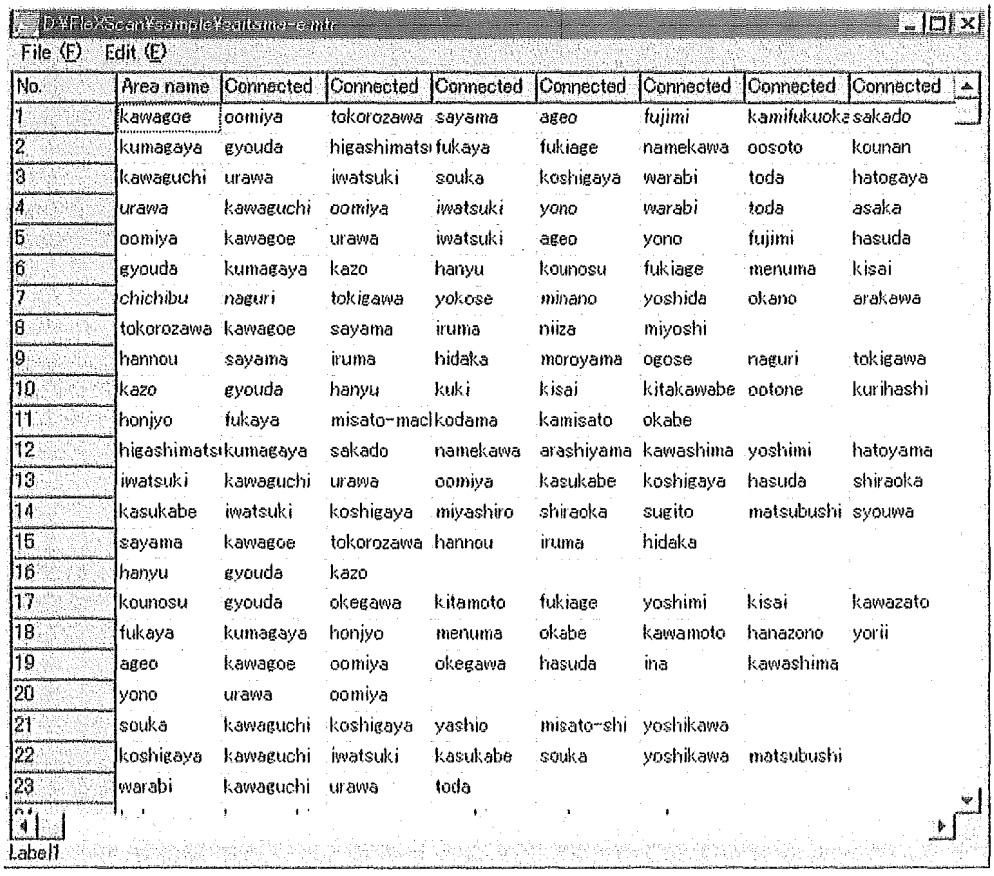


図1　FleXScanの動作画面



No.	Area name	Connected	Connected	Connected	Connected	Connected	Connected	Connected
1	kawagoe	oomiya	tokorozawa	sayama	ageo	fujimi	kamifukuoka	sakado
2	kumagaya	gyouda	higashimatsu	fukaya	fukiage	namekawa	oosoto	kounan
3	kawaguchi	urawa	iwatsuki	souka	koshigaya	warabi	toda	hatogaya
4	urawa	kawaguchi	oomiya	iwatsuki	yono	warabi	toda	asaka
5	oomiya	kawagoe	urawa	iwatsuki	ageo	yono	fujimi	hasuda
6	gyouda	kumagaya	kazo	hanyu	kounosu	fukiage	menuma	kisai
7	chichibu	naguri	tokigawa	yokose	minano	yoshida	okano	arakawa
8	tokorozawa	kawagoe	sayama	iruma	niiza	miyoshi		
9	hannou	sayama	iruma	hidaka	moroyama	ogose	naguri	tokigawa
10	kazo	gyouda	hanyu	kuki	kisai	kitakawabe	ootone	kurihashi
11	honjyo	fukaya	misato-mach	kodama	kamisato	okabe		
12	higashimatsu	kumagaya	sakado	namekawa	arashiyama	kawashima	yoshimi	hatoyama
13	iwatsuki	kawaguchi	urawa	oomiya	kasukabe	koshigaya	hasuda	shiraoka
14	kasukabe	iwatsuki	koshigaya	miyashiro	shiraoka	sugito	matsubushi	syouwa
15	sayama	kawagoe	tokorozawa	hannou	iruma	hidaka		
16	hanyu	gyouda	kazo					
17	kounosu	gyouda	okegawa	kitamoto	fukiage	yoshimi	kisai	kawazato
18	fukaya	kumagaya	honjyo	menuma	okabe	kawamoto	hanazono	yorii
19	ageo	kawagoe	oomiya	okegawa	hasuda	ina	kawashima	
20	yono	urawa	oomiya					
21	souka	kawaguchi	koshigaya	yashio	misato-shi	yoshikawa		
22	koshigaya	kawaguchi	iwatsuki	kasukabe	souka	yoshikawa	matsubushi	
23	warabi	kawaguchi	urawa	toda				

図 2　FleXScan による地区の隣接情報の入力画面

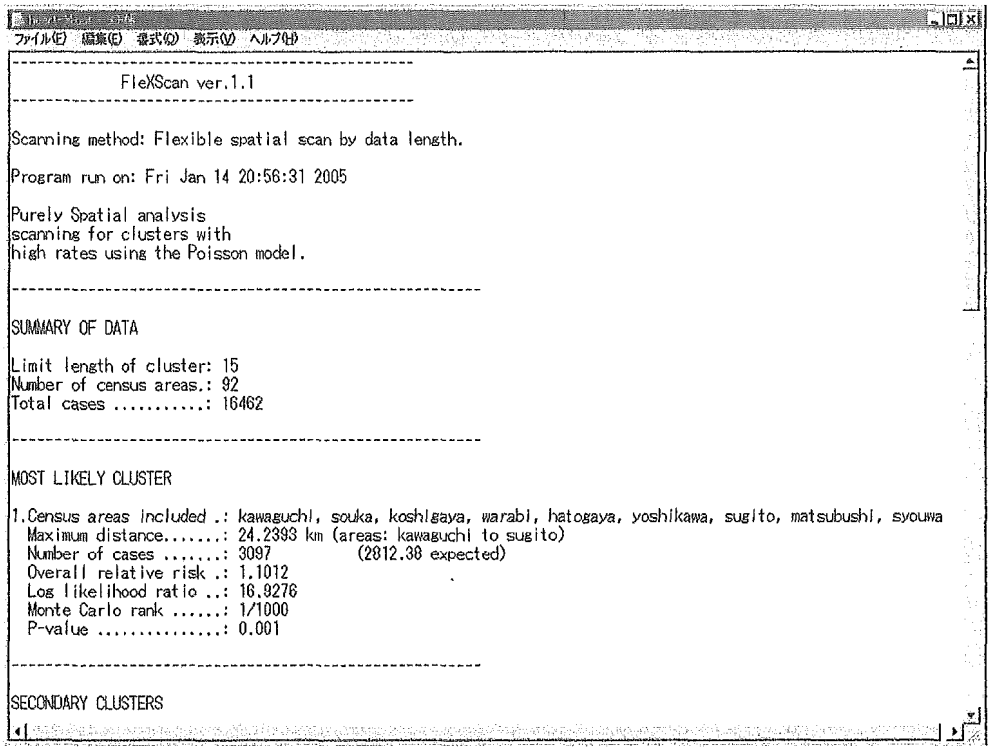


図 3　FleXScan の出力画面(1)：詳細な数値情報が示される

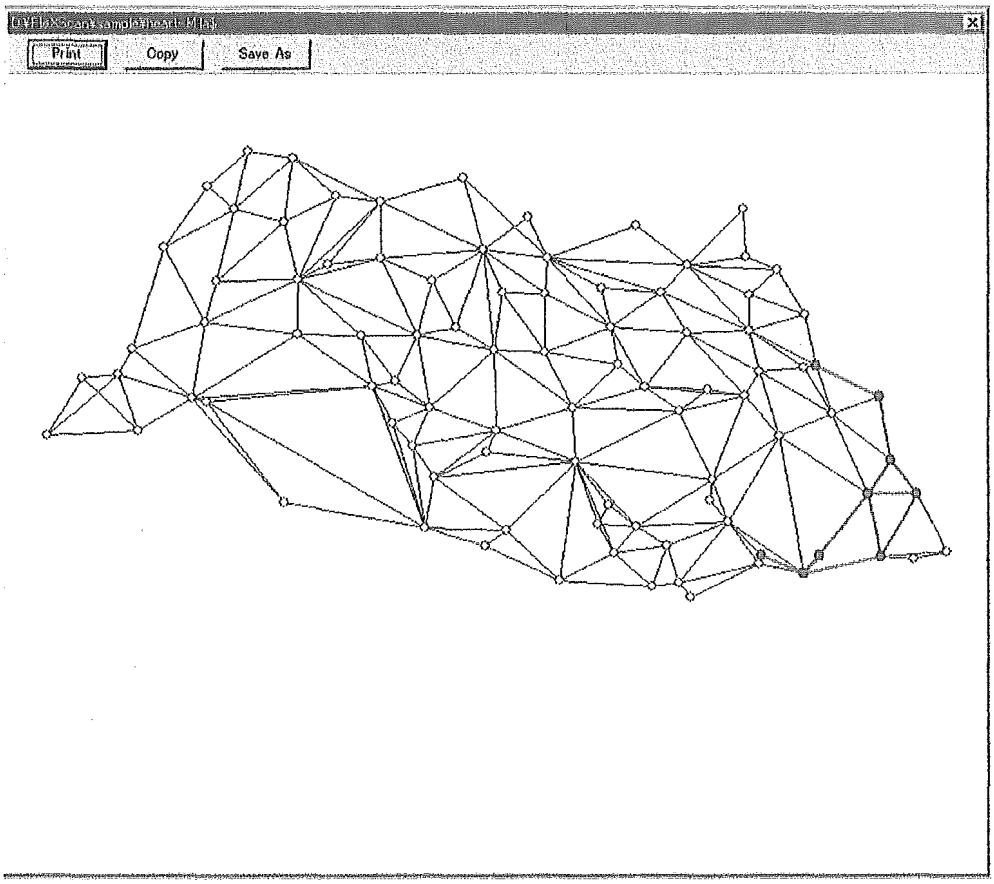


図4　FleXScanの出力画面(2)：対象地域と集積地域の位置が簡略された地図として出力される

厚生労働科学研究費補助金（健康科学総合研究事業）
分担研究報告書

時間的推移に伴うパターン発見のための空間解析手法の開発に関する研究

主任研究者　浅見 泰司　東京大学空間情報科学研究センター教授
協力研究者　片岡 裕介　東京大学大学院新領域創成科学研究科博士課程

研究要旨
本研究の目的は，疾病空間情報の早期の把握，および現象の将来的予測を可能にするための，空間解析手法の開発をおこなうことにある．そのために，感染症の流行状況を対象として，以下のように研究を実施した．(1)定点報告数の母数の推定方法の提案では，1 シーズンの報告数の総数に着目することによって，最尤法を用いた推定をおこなった．(2)疾病空間情報の伝播パターンの分析は，各医療圏の感染者の時系列的な変化を表現する，空間モデルの定式化に関する検討である．(3)地域空間分析のための社会状況の把握においては，インフルエンザの流行と関連性が高い学級閉鎖状況について，位置情報も含めた考察をおこなった．

A．研究目的

現代の我々の生活を取り巻いているのは，高度に複雑化，そして多様化した社会的環境である．そのような現況にありつつも，健康被害状況を迅速かつ的確に把握することは，短期間に有効な意思決定を行うという行政対応を実現するうえで，特に重要視されるべきといえる．そして，その具体的で有効な手段として挙げられるのが，時空間的に感度・精度の高い地域診断システムの構築である．

本研究では，空間情報の時系列データ群のなかから，その空間的特徴を早期に発見するための方法として，空間マイニング・空間推論などの技術に基づいた，空間解析支援ツールの技術開発を行うことを目的とする．

B．研究方法

本研究では，(1)定点報告数の母数の推定方法の提案，(2)疾病空間情報の伝播パターンの分析，および(3)地域空間分析のための社会状況の把握，について検討をおこなった[i]．分析する疾病として，インフルエンザを対象とし，用いる空間単位は二次医療圏とした．以下に，研究方法について示す．

(1) 定点報告数の母数の推定方法の提案

この分析は，最尤法を用いた各医療圏における報告数の母数の推定をおこなうものである．

まず，実際に患者となる確率が医療圏によらず一定とすると，各シーズンの報告数の総数は二項分布に従うとみなせる．また，医療圏数が十分大きいことから，これを正規分布で近似したものの尤度関数を考えることによって，潜在的な感染者数の推定をおこなう方法を提案した．

(2) 疾病空間情報の伝播パターンの分析

ここでは，インフルエンザの流行過程を対象として，現象の将来予測を可能にするための，時系列的な伝播パターンを発見する方法について検討した．

まず，感染の流行過程をあらわすモデルを定式化するにあたり，同一シーズンにおいては，一度罹患すると再度罹患しない，さらに，感染する度合いは，人どうしの接触の度合いに依存する，という前提をたてた．

各医療圏の感染者数を推定するモデルとして，重力モデルを適用し，地域どうしに

方向性の偏りがなく，かつ等方的に影響を及ぼすという条件下で，以前になされた研究で既に整備されているデータを当てはめ，誤差を最小化するようにパラメータを推定した．

(3) 地域空間分析のための社会状況の把握

インフルエンザの流行に関連性の高い社会現象である学級閉鎖状況についての把握をおこなった[i]．

本分析では，2000年 1月 17日から 2月 25日までの，北九州市内の公立小学校の学級閉鎖状況を例として，まず，市内の小学校について住所をもとに位置を特定し，ある時点での閉鎖学級数をあらわす地図を作成した．また，閉鎖学級数を週単位に集計したものと，定点観測データとの比較をおこなった．

（倫理面への配慮）

本研究では集計情報をもとに分析しており，プライバシーにかかわる個人情報は扱っていない．また，疫学倫理指針に基づき研究を行っている．

C．研究成果

(1) 定点報告数の母数の推定方法

報告数を確率密度で表した場合における，最尤法を用いた母数の推計方法を提案し，罹患率と定点観測データの母数との関係式を解析的に導出した．

(2) 疾病空間情報の伝播パターンの分析

重力モデルを応用した，時間的および空間的な感染状況についての定式化をおこなった．これに基づき，360 地域の計量が同時に可能な状態で，パラメータの推定を行い，求められる値の安定性について検証した．

(3) 地域空間分析のための社会状況の把握

住所情報から位置情報を取得し，GIS 上で学級閉鎖状況を把握することが可能な地図を作成した．

また，閉鎖学級数と定点報告数との比較により，学級閉鎖の開始時期から急増する期間において，感染者数が沈静化するという傾向を実証した．

D．考察、E．結論

(1) 定点報告数の母数の推定方法

健康危機情報について，その空間的側面から地域分析を試みる際に，分析の基礎となる各集計データは，何らかの基準化されていることが不可欠である．

本研究では，インフルエンザの流行を地域相互の空間的な関係に着目して検討している．よってここでは，各医療圏の定点による報告数について何らかの基準化が必要となる．

この状況をふまえて，当分析では，定点数が不明な状況下での，最尤法を用いた医療施設の圏域人口の推計方法についての提案を行った．

まず，平均値が m，分散が v の場合の正規分布の確率密度を $f(x;m,v)$ とする．また，j=1...T が時点のパラメータ，i=1...M が医療圏番号とすると，各シーズンの報告数の総数が $x_{(i,j)}$ と表される．

ここで，医療圏 i における潜在患者数が n_i として，そのうち実際に患者となる確率を p（医療圏によらず一定）とすると，$x_{(i,j)}$ は平均値 $n_i p$，分散 $n_i p(1\text{-}p)$ の二項分布に従うとみなせる．また，M は十分大きいので，正規分布で近似できるとすると，$f(x_{(i,j)}; n_i p, n_i p(1\text{-}p))$ が確率密度となる．

そこで，尤度関数 L を求めると，

$$L=\prod_i^M \prod_j^T \left\{ f\left(x_{(i,j)}; n_i p, n_i p(1-p)\right)\right\} \quad (1)$$

となる．

この対数をとった $\ln L$ を最大にする条件から以下の 2 式が与えられる．

$$\begin{cases} \dfrac{\partial \ln L}{\partial n_i}=0 & (2) \\ \dfrac{\partial \ln L}{\partial p}=0 & (3) \end{cases}$$

式(2),(3)から，解析的に，あるいは近似的に関係式が導かれ，これらを連立させることで，(n_i,p)の最尤値が求められる．

ここで，式(2)より与えられるn_iとpの関係式は以下のように解析的に得られる．

$$n_i = \frac{-(1-p)\sqrt{T} + \sqrt{(1-p)^2 T + 4\sum_j^T x_{i,j}^{\ 2}}}{2p\sqrt{T}} \qquad (4)$$

下の図1はT=1，$x_{(i,j)}$=100としたときの(p,n_i)のグラフである．

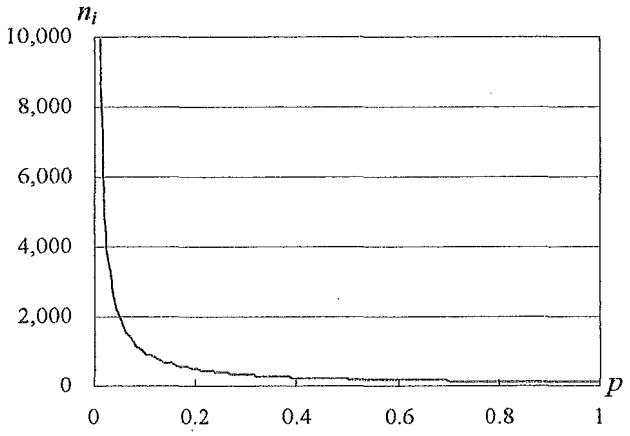


図1　報告数の母数と感染率

さらに，式(4)と，式(3)より得られる近似式から求まる値を用いることによって，空間的なアプローチによる分析上の問題を解決できるのみならず，感染の拡大状況を視覚的に把握できる地域情報地図を作成する上でも有効となる．

(2) 疾病空間情報の伝播パターンの分析

インフルエンザの感染者数を，より適切に予測するために，地域間の流行の方向性に着目したモデルについての検討を行った．

本分析では，360ヶ所にも及ぶ地域を同時に扱うことを想定していることから，まず出来るだけ単純化されたモデルを当てはめた場合について考えてみることにした．

そこで，ある時点における地域どうしの影響を把握するために，重力モデルにもとづいた感染者数を推定するモデルで現実のデータを当てはめてみた．

まず，単純化するために，地域どうしに方向性の偏りがなく，かつ等方的に影響を及ぼすという条件を設定する．

$$\begin{cases} x_{(i,t+1)} = \sum_j \left(\alpha \dfrac{X_{(i,t)} \cdot x_{(j,t)}}{d_{(i,j)}^{\ \beta}} \right) + \varepsilon_i & (5) \\ X_{(i,t+1)} = X_{(i,t)} - x_{(i,t)} & (6) \end{cases}$$

（α, β：重力モデルにおけるパラメータ，$x_{(i,t)}$：時刻tにおける医療施設圏iの感染者数，$X_{(i,t)}$：潜在的な感染者数，$d_{(i,j)}$：医療圏間の距離，ε_i：誤差項）

ここで，誤差を最小化するように，パラメータα，βを推定したところ，図2のように各地域におけるαとβの値が求まった（用いたデータは1999/2000シーズン）．

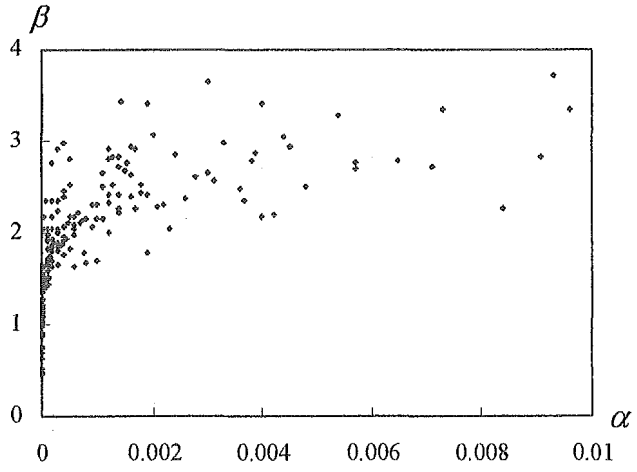


図2　パラメータの推定結果

推定結果を見る限りでは，パラメータの値は安定しているとは言えないが，3シーズン分のデータを用いることで，安定性が上がることが予想される．

モデル化については今後においても，より正確に状況を反映し，予測を早い段階で適切に行うことが可能なモデルに改良する必要があるといえる．

今後は，本研究の成果として得られた定点報告数の母数の推定方法を用いることによって，精度の高い分析が可能となること